Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 11455315; Cache Slough at South Liberty Island near Rio Vista, California

This model archive summary describes the suspended-sediment concentration (SSC) model developed to compute a 15-minute SSC time-series for the period of record: December 20, 2010 to July 09, 2013. This is the first suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in the Office of Surface Water/Office of Water Quality Technical Memorandum and USGS Techniques and Methods, book 3 chapter 4 (USGS, 2016; Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS, 2014).

Site and Model Information

Site number: 11455315
Site name: Cache Slough at South Liberty Island near Rio Vista, California (LIB)
Location: Latitude 38°14'34.84", Longitude 121°41'03.43" referenced to North American Datum of 1983, Solano County, California. Hydrologic Unit 18020163
Equipment: A YSI 6-series sonde began logging turbidity with a model 6136 sensor on December 20, 2010 and was removed on July 16, 2013.

Model number: 11455315.SSC.WY11.1
Model calibration data period: February 9, 2011 – June 28, 2013
Model application date: December 20, 2010 – July 9, 2013
Computed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)
Reviewed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)

Physical Sampling Details and Sediment Data

All sediment data were collected using USGS protocols and are stored in the National Water Information System (NWIS) database: https://waterdata.usgs.gov/nwis (USGS, 2006). Discrete, boat-based samples were collected seasonally, spanning the range of site conditions and specifically targeting large sediment transport events.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). The Equal Discharge Increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge-weighted suspended-sediment samples were collected. Due to the tidal nature of the site, the EDI method was used to collect discharge-weighted samples because velocities are not always isokinetic (from Table 4-5 of TWRI09A4; USGS, 2006). A boat-based discharge measurement was collected immediately before sampling with an Acoustic Doppler Current Profiler (ADCP) to determine the location of each vertical. A FISP US D-96 bag sampler was used to collect the depth-integrated samples at each vertical. The channel cross section can approach a depth of

21 feet in the thalweg with a mean depth of 9 feet. Index velocity data were not collected during the model calibration data period, so the range of velocities at the site is not known. Sediment at this station is mostly fines (93% fines on average) and any potential sampling bias due to non-isokinetic sampling is considered minimal.

Samples collected before January 2012 were analyzed for SSC (mg/L) by the filtration method at the USGS Sediment Laboratory in Marina, California, while those collected after January 2012 were analyzed for SSC by the USGS Sediment Laboratory at its current location in Santa Cruz, California. Though only two samples were analyzed for the percentage of fines (<0.063 mm) during this period, the sand-fine break analysis shows that sediment at LIB contains mostly fines. EDI verticals were analyzed individually for quality control purposes and because sampling occurs in rapidly changing, tidal conditions. Average SSC was computed from the five verticals of each sample for use in the calibration dataset. EDI sets from the same day are not always considered replicates; same day sets are considered unique unless average sample times are less than 45 minutes between sets. Sediment results are publicly available on NWIS and all sediment data were reviewed and approved in the USGS NWIS Water-Quality System database (QWDATA) before being applied in the calibration model.

Surrogate Data

Continuous, 15-minute turbidity data, reported in Formazin Nephelometric Turbidity Units (FNU), were evaluated as an explanatory variable for SSC. Turbidity timeseries data were collected by the USGS California Water Science Center and were analyzed and approved per USGS guidelines (Wagner and others, 2006). These data are located at: https://waterdata.usgs.gov/ca/nwis/uv?site_no=11455315. Discharge was not considered as a second explanatory variable because concurrent index velocity data were not collected during the waterquality monitor deployment.

Model Calibration Dataset

The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair surrogate data with discrete sediment data (Domanski and others, 2015). Turbidity was paired to each sediment sample with a matching window of ± 15 minutes. The SAID manual is available at: https://pubs.er.usgs.gov/publication/ofr20151177.

Four EDI sets (A, B, C, D) were collected on 11/29/2011. The set averages are over 45 minutes apart, except for Set D (Set A = 1239 pst, Set B = 1335 pst, Set C = 1441 pst, Set D = 1512 pst). Sets A, B and D are included as individual samples in the calibration dataset. Set C was removed from the dataset to increase the amount of time between the second and third sets.

Two EDI sets (A and B) were collected on 02/24/2012. The sampling time span for the two sets exceeds one hour and the set averages are over 45 minutes apart (Set A = 1013 pst, Set B = 1306 pst). Both sets were included as individual samples in the calibration dataset.

Two EDI sets (A and B) were collected at the site on 03/15/2012. The sampling time span for the two sets exceeds one hour and the set averages are over 45 minutes apart

(Set A = 1120 pst, Set B = 1206 pst). Both sets were included as individual samples in the calibration dataset.

The final calibration dataset is compiled from 17 concurrent measurements of SSC and turbidity. Summary statistics and the complete model calibration dataset are provided in the following sections.

Model Development

Two simple linear regression (SLR) models were evaluated: Model 1) linear model with one explanatory variable (turbidity) and Model 2) $\log_{10}$-transformed model with one explanatory variable (turbidity).

Diagnostic statistics and plots for model review were output using a combination of Matlab, SAID, and the R environment (R Core Team, 2018). The regression methods used are described in Helsel and Hirsch (2002). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate regression models. The best model was chosen based on residual plots, root-mean-squared error (RMSE), coefficient of determination (R^2), significance tests (p-values) and prediction error sum of squares (PRESS) statistics. RMSE and PRESS statistics cannot be used to compare regressions with different response variable units, so R^2, Mean Square Prediction Error (MSPE) values and residual plots were used as the main determinants of model strength. Values for these statistics were computed for two models and are included below. The best SLR model is a log model with $\log_{10}$-transformed turbidity as the surrogate (Model 2, highlighted in the table below).

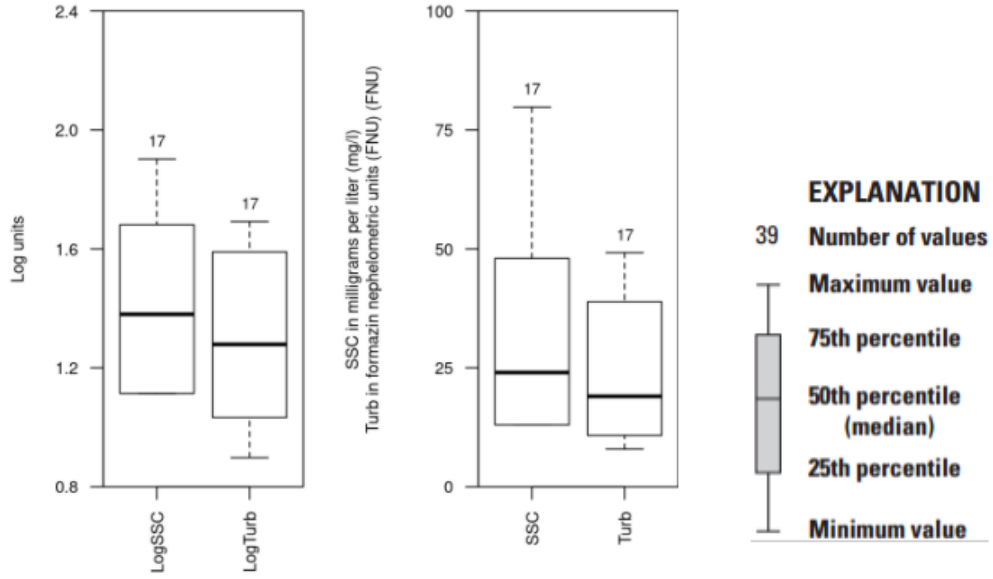
No.	R^2	R^2_a	RMSE	PRESS	MSPE	n	Type
Model 1	0.81	0.79	9.85	1929	30.3	17	Linear
Model 2	0.91	0.91	0.09	0.15	20.4	17	Log

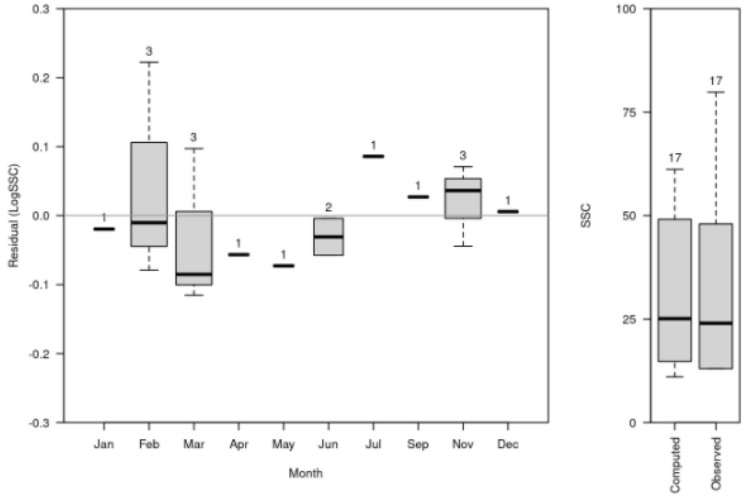
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals were inspected for values greater than three or less than negative three; values outside this range are considered potential extreme outliers. The studentized residuals were reviewed from the output reports and none of the samples were deemed to be extreme outliers. All 17 observations were retained in the model.

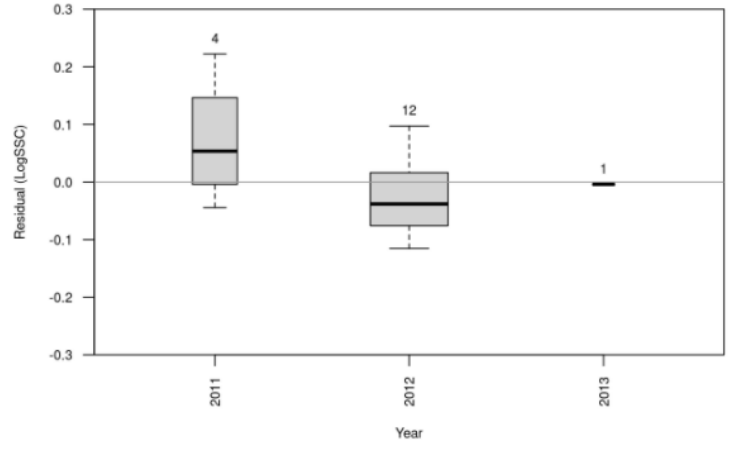
Plots

The following plots were generated using a R-based application (Version 1.0) developed by Patrick Eslick of the USGS Kansas Water Science Center, which is available at: http://kswsc.cr.usgs.gov:3838/peslick/ModelArchiveSummary/.

Boxplots of turbidity and SSC data show the range of measured data for each parameter. The second set of boxplots show SSC residuals by month and water year.

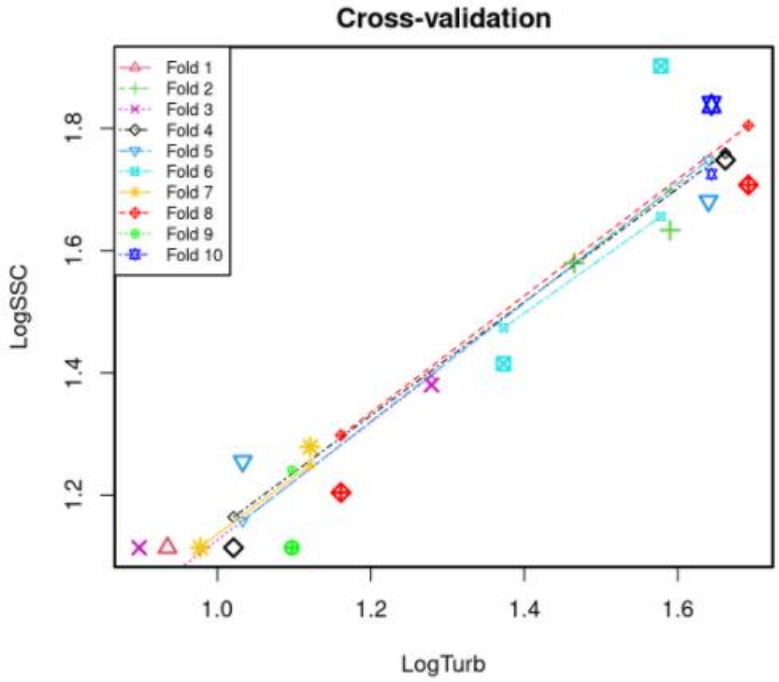

Log units
2.4
2.0
1.6
1.2
0.8
17
17
LogSSC
LogTurb
SSC in milligrams per liter (mg/l)
Turb in formazin nephelometric units (FNU) (FNU)
100
75
50
25
0
17
17
SSC
Turb
EXPLANATION
39 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value


Residual (LogSSC)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
3
3
1
1
2
1
1
3
1
Jan
Feb
Mar
Apr
May
Jun
Jul
Sep
Nov
Dec
Month
SSC
100
75
50
25
0
17
17
Computed
Observed

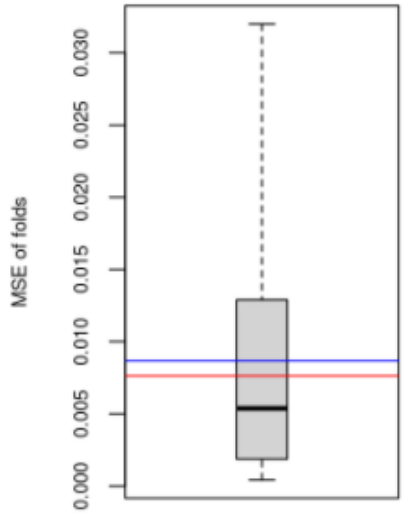

Residual (LogSSC)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
4
12
1
2011
2012
2013
Year

Cross Validation

The cross-validation graph below shows a k-fold validation with k=10. The points represent observations that were left out of each fold.



Minimum MSE of folds:	4.29E-04
Mean MSE of folds:	0.01
Median MSE of folds:	0.01
Maximum MSE of folds:	0.03
(Mean MSE of folds) / (Model MSE):	1.14



Red line - Model MSE
Blue line - Mean MSE of folds

Model Summary

The final SSC model at LIB is a SLR model based on 17 concurrent measurements of SSC and turbidity data collected over three water years. The model is shown below with basic model information, regression coefficients, correlation and summary statistics:

Linear Regression Model	Coefficient of Determination (R^2)
$log10SSC = 0.202 + 0.937 * log10Turb$	0.91

where
SSC = suspended-sediment concentration, in milligrams per liter (mg/L) and
Turb = turbidity, in formazin nephelometric units

SSC was transformed during model development, so the computed prediction may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF), shown below.

Model	Start date	End date	Linear Regression Model	*BCF*
1	12/20/2010	07/09/2013	$SSC = 10^{0.202} \times Turb^{0.937} \times BCF$	1.02

The SSC time-series is computed from USGS turbidity data. Minimum and maximum turbidity values for the model application period are listed below. SSC time-series data exceeding extrapolation limits were removed. This model cannot be used to extrapolate more than 10% above or below the range of samples in the calibration dataset (USGS, 2016). The extrapolated, maximum computed SSC for this model is 88 mg/L. The original maximum, computed SSC was 447 mg/L.

Parameter	Minimum	Maximum
Computed SSC (mg/L)	0	88
Turbidity (FNU)	0	396

Suspended-Sediment Concentration Record

The SSC record computed using this regression model is stored in the USGS National Real-Time Water Quality (NRTWQ) website. The complete record can be found at: http://nrtwq.usgs.gov/ca.

Model

Log10SSC = 0.202 + 0.937Log10Turb

Variable Summary Statistics

	Turb	log10Turb	SSC	log10SSC
Minimum	7.9	0.9	13	1.11
1st Quartile	10.7	1.0	13	1.11
Median	19	1.3	24	1.38
Mean	24.6	1.3	32.5	1.42
3rd Quartile	40.1	1.6	49	1.69
Maximum	49.2	1.69	80	1.90

Basic Model Statistics

Number of observations	17
Root Mean Squared Error (RMSE)	0.09
Model Standard Percentage Error (MSPE)	20.4
Coefficient of determination (R^2)	0.91
Adjusted R^2	0.91
Bias Correction Factor	1.02

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.202	0.10	2.01	6.25E-02
log10Turb	0.937	0.08	12.44	2.63E-09

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.977
E.vars	-0.977	1.000

Outlier Test Criteria

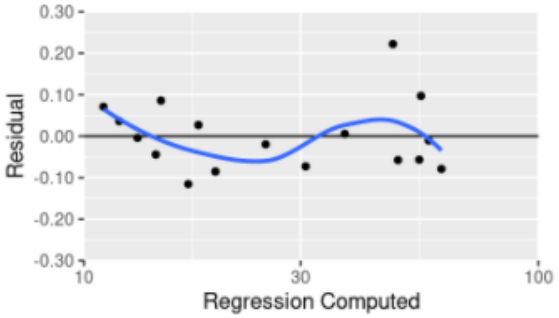
Leverage	Cook's D	DFFITS
0.353	0.192	0.686

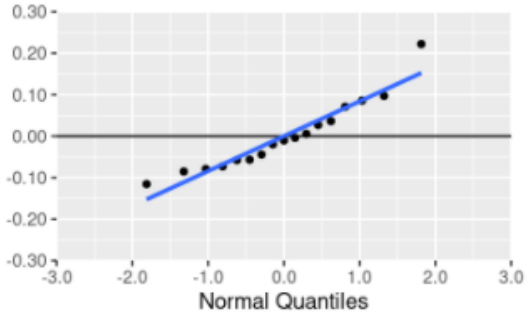
Flagged Observations

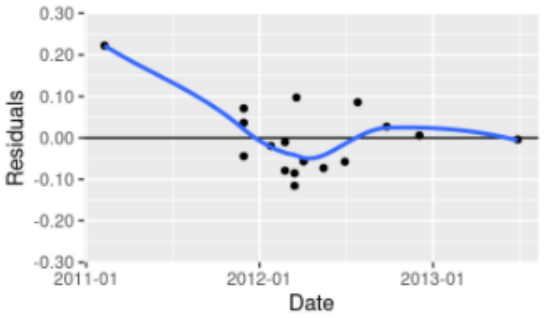
Date	Time	LogSSC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
2/9/2011	14:35	1.9	1.68	0.222	2.7	3.65	0.114	0.472	1.31

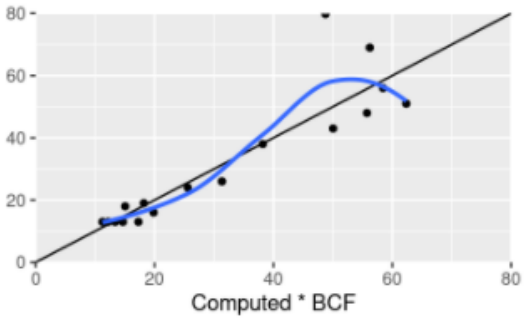
Residual diagnostic plots

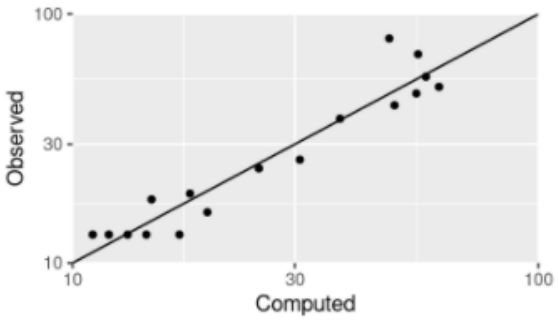
Plots were generated using the model archive summary application developed by Patrick Eslick of the USGS Kansas Water Science Center.

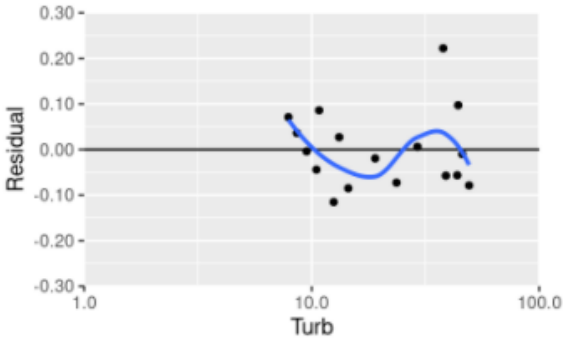












Model-Calibration Dataset

	Date & Time	SSC	log10SSC	Turb	log10Turb	Computed LogSSC	Computed SSC	Residual	Normal Quantile
1	2/9/2011 14:35	80	1.90	37.8	1.58	1.68	48.74	0.224	1.81
2	11/29/2011 12:39	13	1.11	7.9	0.90	1.04	11.25	0.071	0.81
3	11/29/2011 13:35	13	1.11	10.5	1.02	1.16	14.68	-0.045	-0.30
4	11/29/2011 15:12	13	1.11	8.6	0.93	1.08	12.18	0.037	0.62
5	1/25/2012 10:37	24	1.38	19	1.28	1.40	25.59	-0.020	-0.15
6	2/24/2012 10:13	51	1.71	49.2	1.69	1.79	62.38	-0.079	-1.03
7	2/24/2012 13:06	56	1.75	45.9	1.66	1.76	58.45	-0.010	0.00
8	3/15/2012 11:20	16	1.20	14.5	1.16	1.29	19.87	-0.086	-1.32
9	3/15/2012 12:06	13	1.11	12.5	1.10	1.23	17.29	-0.115	-1.81
10	3/19/2012 13:54	69	1.84	44.1	1.64	1.74	56.30	0.097	1.32
11	4/3/2012 11:17	48	1.68	43.6	1.64	1.74	55.71	-0.056	-0.45
12	5/15/2012 12:08	26	1.41	23.6	1.37	1.49	31.35	-0.073	-0.81
13	6/29/2012 10:16	43	1.63	38.9	1.59	1.69	50.06	-0.058	-0.62
14	7/26/2012 10:44	18	1.26	10.8	1.03	1.17	15.08	0.085	1.03
15	9/25/2012 10:16	19	1.28	13.2	1.12	1.25	18.19	0.027	0.45
16	12/3/2012 10:32	38	1.58	29.2	1.47	1.57	38.27	0.005	0.30
17	6/28/2013 11:16	13	1.11	9.5	0.98	1.12	13.37	-0.004	0.15

Definitions

SSC: Suspended sediment concentration (SSC) in mg/l (80154)
Turb: Turbidity in FNU (63680)

References

Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., https://pubs.usgs.gov/of/2015/1177/ofr20151177.pdf.

Duan, Naihua, 1983, Smearing estimate – A nonparametric retransformation method: Journal of the American Statistical Association, Volume 78-383, 605-610 p.

Edwards, T.K. and Glysson, G.D., 1999, Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. C2, 89 p., Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

R Core Team, 2018, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available from: https://www.R-project.org/.

Rasmussen, P.P., Gray J.R., Glysson G.D., Ziegler A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods book 3, chap. C4, 53 p., Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations. book 9, chap. A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf.

[USGS] U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Oblinger CJ, Smith BA, 2006, Guidelines and standard procedures for continuous water quality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3, Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.